Science
AAAS

Supplementary Materials for

Transcription factor AP2 controls cnidarian germ cell induction

Timothy Q DuBuc, Christine E Schnitzler, Eleni Chrysostomou, Emma T McMahon, Febrimarsa, James M Gahan, Tara Buggie, Sebastian G Gornik, Shirley Hanley, Sofia N Barreira, Paul Gonzalez, Andreas D Baxevanis, Uri Frank

Correspondence to: uri.frank@nuigalway.ie

This PDF file includes:

Materials and Methods
Figs. S1 to S12
Captions for Table S1
Captions for Files S1 to S2
Captions for Movies S1 to S2
References 41-63

Other Supplementary Materials for this manuscript include the following:

Data S1 to S2
Movies S1 to S2

Materials and Methods

Animal husbandry

Adult *Hydractinia symbiolongicarpus*, male (291-10) and female (295-8) strains, were maintained in cycling aquaria in artificial seawater (salinity 35 ppt). Animals were grown on glass slides in staining racks and were fed four times a week with *Artemia franciscana,* and once a week ground oyster. Fertilized eggs were collected by exposing animals to constant darkness for 10 hours, followed by light exposure. Males and females spawn 1.5 hours after exposure to light.

Micromanipulation of embryos

<u>Microinjection setup</u>
Fertilized eggs were transferred to a Petri dish coated with 200 micron Nitex mesh screen. Zygotes are approximately 180 microns and settled in the holes. Cells were injected, prior to first cleavage, using a Narishige IM 300 microinjection system. To delay cleavage, zygotes were stored at 4°C prior to injection.

<u>Stable transgenic reporters</u>
Zygotes were first injected with plasmid DNA at a concentration between 500-1000 ng/µl. Injected plasmid was supplemented with 100 mM of KCl, which helps reduce mosaic integration. Codon optimized mScarlet fluorescent protein (*41*) was synthesized by Eurofins Genomics. All injected animals were raised through larval development and metamorphosed on glass slides by a 3 hour 100 mM CsCl pulse treatment (*42*). Animals were bred to produce stable, non-mosaic transgenic offspring.

<u>CRISPR/Cas9 mutagenesis</u>
Single guide RNAs (sgRNAs) were predicted using Geneious (2017.9.1.8). The two selected sgRNAs did not match other genomic sequence. sgRNAs were synthesized by Integrated DNA Technologies (IDT) and were hybridized with tracrRNA (IDT, Cat. #1072532) as instructed. Hybridized sgRNAs were incubated with recombinant Cas9 (IDT, Cat. #1074181) for fifteen minutes prior to being microinjected into zygotes. The final concentration of injected material was 1µg/µl Cas9 and 500 ng/µl of sgRNA+tracrRNA. Methods for genotyping are explained below.

<u>Ectopic expression</u>
Ectopic expression of *Tfap2* was achieved by microinjection of one of three different plasmids (fig. S10A to C). Plasmids were injected at 500-1000 ng/µl and were supplemented with 100 mM of KCl. Animals were scored at 24 or 48 hours post fertilization and after metamorphosis.

shRNA

A delay of ectopic expression was achieved by co-injection of plasmid and a short hairpin RNA (shRNA) targeting GFP (shGFP) sequence of 5'GATGACGGGAACTACAAGACA3'.

The consensus sequence of 5' TAATACGACTCACTATA**GGATGACGCGATCTG-CAAGACA**~~ATTTACT~~TGTCTTGTAGTTCCCGTCATCTT 3' was synthesized by IDT and contains the T7 promoter, passenger (bold) and loop sequence (underlined). shGFP was synthesized as previously established (*37*) and was quantified using a Nanodrop. shGFP was microinjected at 500 ng/µl.

Animals that were co-injected with plasmid and shRNA recovered GFP expression approximately ten days post injection (fig. S12). Delayed ectopic expression experiments were conducted across three consecutive days and animals were collected at days 8-10. Experiments were carried out three times and consisted of a total of nine separate injections.

Genotyping

Sixty CRISPR-Cas9 injected animals were metamorphosed. A subset of 32 animals were healthy. Of these, nine animals displayed defects in sexual development. Five animals were analyzed for *Tfap2* mutations. Primers spanning the entire coding region of the gene (Fwd1 – TGAAAGGTCAAGAATCCATTGCG, Rev2 – TGATTTTGTTGATGAACGCTGTCG) were used in a PCR reaction to identify large deletions. Two additional primer pairs spanning exon 1 (Fwd1, Rev1 – CGCACGAGTTGGTTTACTTTGAATG) or exons 2 and 3 (Fwd2 – TCCTTCCTGTTTGTCTAATCCCGTC, Rev2) were used to identify mutations caused by a single sgRNA. PCR products were cloned (see below) and sequenced using the DNA specific primers listed above. For the majority of samples, the Fwd1 primer was used for sequencing and provided enough sequencing coverage for us to identify mutations.

DNA was extracted using the Qiagen DNAeasy DNA extraction kit (Cat. 69504, Qiagen). Tissue samples consisted of five polyps that were removed from a growing colony. The only changes to the manufacturers protocol we made was in the tissue preparation. Polyps were centrifuged (3000g for 5 seconds) prior to lysis to remove excess seawater, then processed following the manufacturer's guidelines.

A two-step PCR approach was used to amplify genomic fragments. PCR conditions using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Cat. #F530S,) were as follows: initial denaturation (98°C, 3:00), denaturation (98°C, 0:30), annealing/extension (68°C, 1:15), repeat steps 2-3 (30x), and a final extension (68°C, 10:00). PCR reactions were run on a 1% agarose gel and positive bands were excised and purified using a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Cat. #740609.10). A-overhangs were added by conducting an additional round of PCR using MyTaq™ (Cat. Bio-21105) and fragments were subsequently ligated into the pGEM®-T Easy Vector System (Promega, Cat. #A1360). DH5-alpha *E. coli* bacteria were transformed and plated on ampicillin LB-agar plates and grown overnight. Individual clones were picked and grown in overnight cultures of LB-broth (with ampicillin, 100µg/ml). Plasmid DNA was extracted using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, Cat. #PLN250-1KT). Samples were sequenced and analyzed in Geneious to identify mutations.

Homozygous knockout animals were generated by crossing mutant 4 (a mosaic animal) with a *Tfap2* wild type animal that carried a *Piwi1* reporter transgene. F1 animals were metamorphosed and grown to sexual maturity. One male and one female that displayed defects in sexual polyp phenotypes (reduced number of germ cells) and carried the *Piwi1* reported transgene were genotyped. Each of these two sibling animals is a heterozygote carrying a 5' insertion mutation in one *Tfap2* allele that resulted in a premature stop codon, and had a wild type allele (fig. S8). They were crossed and F2 offspring were genotyped identifying a mutant homozygote animal. We performed genotyping on this animal from two independent DNA extractions and all sequences came back as having an insertion mutation.

Phylogenetic analysis

Tfap2 protein sequences were gathered from diverse taxa using the open access platform (reefgenomics.org) (*43*). Additional proteins were collected from EnsemblMetazoa (EMBL, www.metazoan.ensembl.org/index.html), recent shark genome databases (*44*), the *Clytia* genome (*45*) and GenBank resources (NCBI, https://www.ncbi.nlm.nih.gov/genbank/).

Sequences were aligned using MAFFT (*46*) and were trimmed to only contain the putative DNA binding domain. Trees were constructed using MrBayes software (*47*) consisting of 4,000,000 generations using "mixed" models. Maximum-likelihood tree bootstraps were constructed from 100 replicates and trees were edited in FigTree (version 1.4.0) (*48*).

In situ hybridization

Tissue samples were collected from anaesthetized animals in ice cold 4% $MgCl_2$ (in 50% distilled water / 50% filtered seawater). Animals were fixed in ice cold 0.2% glutaraldehyde, 4% paraformaldehyde in filtered seawater for 90 seconds. A second fixation in 4% paraformaldehyde in filtered seawater was then added and samples were placed at 4°C for one hour.

Fixative was removed and samples were washed with phosphate buffered saline (PBS) supplemented with 0.1% Tween20 (PBST). Tissue was then dehydrated in methanol and stored at -20°C. Before hybridization, tissue samples were permeabilized in 75% acetone in 25% methanol, then rehydrated into PBST.

Probes (either DIG or fluorescein labelled) were hybridized at 58°C and detected using Anti-DIG/fluorescein antibodies (Roche, Cat. # 11093274910, 11426338910, 11207733910, 11426346910). Alkaline phosphatase-conjugated antibodies (AP) were used for colorimetric reactions and the peroxidase detection system (POD) was used for fluorescent reactions according to previously established protocols (*49*).

Disassociation and flow cytometry

Tissue dissociation

Colonies were placed in 4% $MgCl_2$ in 1:1 artificial sea water:diH_20 for relaxation (5-15minutes). The polyps were then removed from the colonies and placed in 0.5% pronase (Sigma-Aldrich, Cat. #10165921001) in filtered sea water (20 polyps/200 µl) for 2 hours with constant rocking at room temperature. Every 30 min a gentle mixing was applied. Once the tissue was fully dissociated, the reaction was stopped by adding BSA to a final concentration of 0.1%. Next, the cell suspension was passed through a 100 µm filter (Sysmex Partec, Cat. #04-004-2328) to remove any residual cell clumps.

Flow Cytometry

Cell suspensions were labelled with 37.5µg/ml Hoechst 33342 (Sigma-Aldrich Cat.#14533) for 20 min at 18°C. Flow cytometric analysis was performed using a BD FACSCanto™ II flow cytometer (BD Biosciences, San Jose, CA, USA) which was calibrated according to the manufacturers' recommendations. Data were analyzed using Diva® v8.0.1 (BD Biosciences) or FlowJo™ (TreeStar Inc., Olten, Switzerlansd). Typical gating strategies involved gating on live, nucleated cells (Hoechst 33342 positively stained population) measured using the 405nm excitation laser line and 450/50nm photomultiplier tube (PMT) fluorescence detector), followed by doublet exclusion based on Forward Scatter Height (FSC-H) versus Forward Scatter Area (FSC-A). Expression of the GFP reporter in transgenic animals was measured using the 488nm excitation laser line in the 530/30nm PMT fluorescence detector. Cells prepared from wild type animals were used as a gating control.

Fluorescence-activated cells sorting (FACS)

For Fluorescence Activated Cell Sorting (FACS) one hundred polyps from each reporter line (Piwi1::GFP, Tfap2::GFP male and Tfap2::GFP female (n=3)) were dissociated as described previously and re-suspended in 1ml of filtered ASW containing 0.5% Pronase and 0.1% BSA. FACS was performed on a BD FACSAriaII™ high speed cell sorter using a 100um nozzle at 20 psi with filtered ASW as sheath. ASW was used as sheath fluid to eliminate osmotic stress to the cells during sorting. GFP expressing and non GFP expressing cell populations were gated on live, nucleated single cells as described in Extended data Fig. 4. Cells prepared from matched wild type animals were used as a gating controls. Cell populations were sorted directly into Trizol (Life Technologies, Cat. #15596026) for subsequent RNA extraction using the Direct-zol™ RNA MiniPrep kit (Zymo Research Cat. #R2050). The purity of the sorted populations was checked by sorting directly into ASW followed by flow cytometric analysis on a BD FACSAriaII™ cell sorter.

Imaging Flow Cytometry

For imaging flow cytometry cell suspensions were prepared, as previously described, from a dual reporter animal: Piwi1::GFP, BetaTub::mScarlet and analyzed using INSPIRE™ software on an Imagestream X Mark II™ imaging flow cytometer (Luminex Corporation, IL, USA).

Analysis was performed using IDEAS™ software (Luminex Corporation, IL, USA). Samples were gated on focused cells (bright-field gradient root mean squared), cells as measured by brightfield area versus aspect ratio followed by Hoechst 33342 positive population and GFP versus mScarlet measured fluorescence fig. S10. Hoechst, GFP and mScarlet fluorescence was measured using the 405nm excitation laser line and 435-505nm bandpass filter; 488nm excitation laser line and 480-560nm bandpass filter and 561nm laser excitation line and 595-640nm bandpass filter respectively. Cells from single reporter animals Piwi1::GFP and BetaTub::mScarlet were used a compensation and as gating controls.

Antibody generation

For Piwi1 antibody production, we designed the following primer sets to amplify a similar region of Piwi1 as previously reported in the cnidarian *Hydra* (*50*) using the following primers that also contained a 5-prime HIS-tag for purification.

Piwi1- Fwd : AAAAAACATATGCACCATCACCATCACCACATGACAGG-TAGAGCGAGAGGAAGATC

Piwi1- Rev : TTATATGGATCCCTATGGAACCAA-ATTAGTGTAAGTAAC

Fragments were cloned and expressed in pET3s expression vector using NDE1 and BamH1 restricted enzymes (encoded in the primer sequence) and expressed in Rosetta DE3 pLysS bacteria. Protein was extracted and analyzed electrophoretically. Piwi1 fragments were injected into rabbits (two animals) by Eurogentec. Antisera were analyzed by western blots and immunofluorescence fig. S2A and B.

RNA Isolation from tissue samples

RNA was collected from polyp tissue collected from wild type *Hydractinia echinata* living on hermit crab shells. Samples consisted of three replicates each of (a) mixed male and female sexual polyp heads, (b) mixed male and female feeding polyp heads, (c) male sexual polyp bodies, (d) female sexual polyp bodies, (e) male feeding polyp bodies, and (f) female feeding polyp bodies. Colonies were first placed in 4% $MgCl_2$ (in 50% distilled water / 50% filtered seawater) to relax animals for approximately five minutes. Polyps were removed from the shell using Vannas curved surgical scissors (Mason Technology, 80mm Cat.#1341). Once removed, samples were transferred to a petri dish (still in $MgCl_2$) and excess debris was removed. Sexual tissue and feeding tissue were separated into two petri dishes. Using a razor blade, head tissue was removed from the body and the samples were sorted into separate 1.5ml Eppendorf tubes. Tissue samples were centrifuged at 6,000g and excess seawater was removed. RNA was isolated using Trizol and was subsequently cleaned the QIAGEN® RNAeasy Mini extraction kit (QIAGEN, Cat.#74104).

Library preparation and sequencing

After RNA isolation, samples were shipped to the NIH Intramural Sequencing Center (NISC) for further processing and sequencing. For the 18 tissue samples described above, RNA-Seq libraries were constructed from 1 μg RNA using the Illumina TruSeq RNA Sample Prep Kit, version 2. The resulting cDNA was fragmented using a Covaris E210 focused ultrasonicator. Library amplification was performed using 10 cycles to minimize the risk of over-amplification. Unique barcode adapters were applied to each library. Libraries were pooled in an equimolar ratio and sequenced together on three lanes of an Illumina HiSeq 2500 sequencing system using version 4 flow cells and sequencing reagents. At least 49 million 125-base read pairs were generated for each individual library. The bulk cell-sorted RNA samples were amplified with the Ovation® RNA-Seq System V2 kit, and sequencing libraries were made with the Illumina TruSeq Stranded mRNA Library Prep Kit. Libraries were pooled and sequenced on two lanes of an Illumina NovaSeq 6000 sequencing system (2x151 bp).

Quality control and trimming

Read quality control for both sets of samples was performed using FastQC v0.11.6. Overrepresented sequences and low-quality bases were trimmed using Trimmomatic v0.36 (*51*). Samples with > 25% adapter sequences (two out of six samples from the *Tfap2* reporter line) were excluded from the analysis. After trimming, unpaired reads and reads shorter than 36 bp were discarded.

Reference transcriptome assembly and annotation

A strand-specific Trinity transcriptome was constructed from adult *H. echinata* tissues (see Data Availability). RNA was isolated from *H. echinata* strain F4 (a female strain), including three samples of pooled feeding and sexual polyps. RNA-Seq libraries were constructed from 1 μg RNA using the Illumina TruSeq Stranded mRNA Library Prep kit and sequenced together on a single lane of an Illumina HiSeq 2500 sequencing system as described above. The transcriptome was assembled using Trinity (v2.1.1) with default parameters, indicating that the libraries were stranded. Coding regions and protein sequences were predicted using TransDecoder (v3.0.1) before annotating the transcriptome using MMseqs2 with default parameters (*52*). This annotated transcriptome was used as the reference transcriptome in analyses of the 18 tissue samples.

Reads from all bulk cell sorted samples were combined to generate two transcriptome assemblies. The first transcriptome was assembled using Trinity (*53*) (v2.8.4) with default parameters. For the second assembly, reads were mapped to a draft assembly of the *H. symbiolongicarpus* genome using HISAT2 (*54*) (v2.1.0). The resulting alignments were sorted using samtools and used to assemble transcripts with Stringtie (v1.3.4). A consensus, non-redundant transcriptome was made by combining the Trinity and Stringtie assemblies with the tr2aacds.pl Perl script from the EvidentialGene package (*55*) (http://arthropods.eugenes.org/EvidentialGene/trassembly.html). Coding regions and protein sequences were predicted using TransDecoder (v5.5.0) before annotating the transcriptome using PANNZER2 (*56*), as well as with a blastx sequence similarity search against the NCBI non-redundant protein database. This annotated transcriptome was used as the reference transcriptome in ensuing analyses of the bulk cell sorted samples.

Read alignment and gene-level read counts

Reads from the 18 tissue samples were aligned to the *H. echinata* strand-specific reference transcriptome using bowtie2 (v2.2.6) and gene-level read counts were generated with RSEM via scripts provided in the Trinity package (align_and-estimate_abundance.pl and abundance_estimates_to_matrix.pl). Overall alignment rates from bowtie2 ranged from 78-84.5% for each sample. The resulting counts.matrix file was used for downstream analyses of differential expression. Reads from each bulk cell sorted sample were aligned back to the non-redundant reference transcriptome using HISAT2 and gene-level read counts were generated with HTSeq-count (*57*) (v0.11.2). Overall alignment rates from HISAT2 ranged from 68% to 87% for each sample.

Differential Expression analysis for tissue samples

For the 18 tissue samples, unwanted variation (such as from batch effects and library preparation) was removed from RNA-seq data sets by performing between-sample normalization methods using the R package RUVseq (*58*). Differential expression analysis between sexual heads and feeding heads, between female sexual bodies and female feeding bodies, between male sexual bodies and male feeding bodies, between male sexual bodies and female sexual bodies, and between male feeding bodies and female feeding bodies was performed with R packages EdgeR (*59*), DEseq2 (*60*), and Sseq (*61*). Results from all tools were intersected and genes were reported as differentially expressed if found by all three tools to be up or downregulated with a false discovery rate (EdgeR) or adjusted p-value (DESeq2, sSeq) < 0.05 and $\log_2$ fold change of >2 or <-2.

Transcriptomic characterization of germ cells, i-cells, and somatic cells

To identify genes expressed in each cell type without using an arbitrary minimum count threshold, we used the method described in Hebenstreit et al. (*62*). We transformed the read count data using the rlog function of DESeq2 and calculated the mean rlog across replicates for each cell type. We then used the normalmixEM function of the mixtools (*63*) R package to fit two Gaussian curves to the bimodal distribution of mean rlog values. As described in Leclère et al. (*45*), we defined expressed ("on") genes as the genes that have a posterior probability > 0.5 to belong to the high expression distribution. We used these data to generate a Venn diagram of genes expressed in germ cells (Tfap2+), i-cells (Piwi1+), and somatic cells (GFP-negative cells form the piwi reporter line). GO term enrichment analysis of cell-type specific genes was performed using the R package topGO (v2.36.0).

Imaging

Images of transgenic animals or whole colonies were taken using a Leica M165FC microscope with a Leica DFC7000T camera. Movies were generated using a Leica MZ16FA microscope with a Leica DFC350FX camera. Movies were edited using Fiji ImageJ (Version 2.0.0-rc-69/1.52n). Photos of in situ hybridization samples were gathered using an Olympus BX15 compound microscope with a Olympus DP72 camera attached by a U-TV0.63xC adaptor. Confocal images were collected using an inverted Olympus Fluoview 1000 laser scanning confocal microscope.

Supplementary Figures

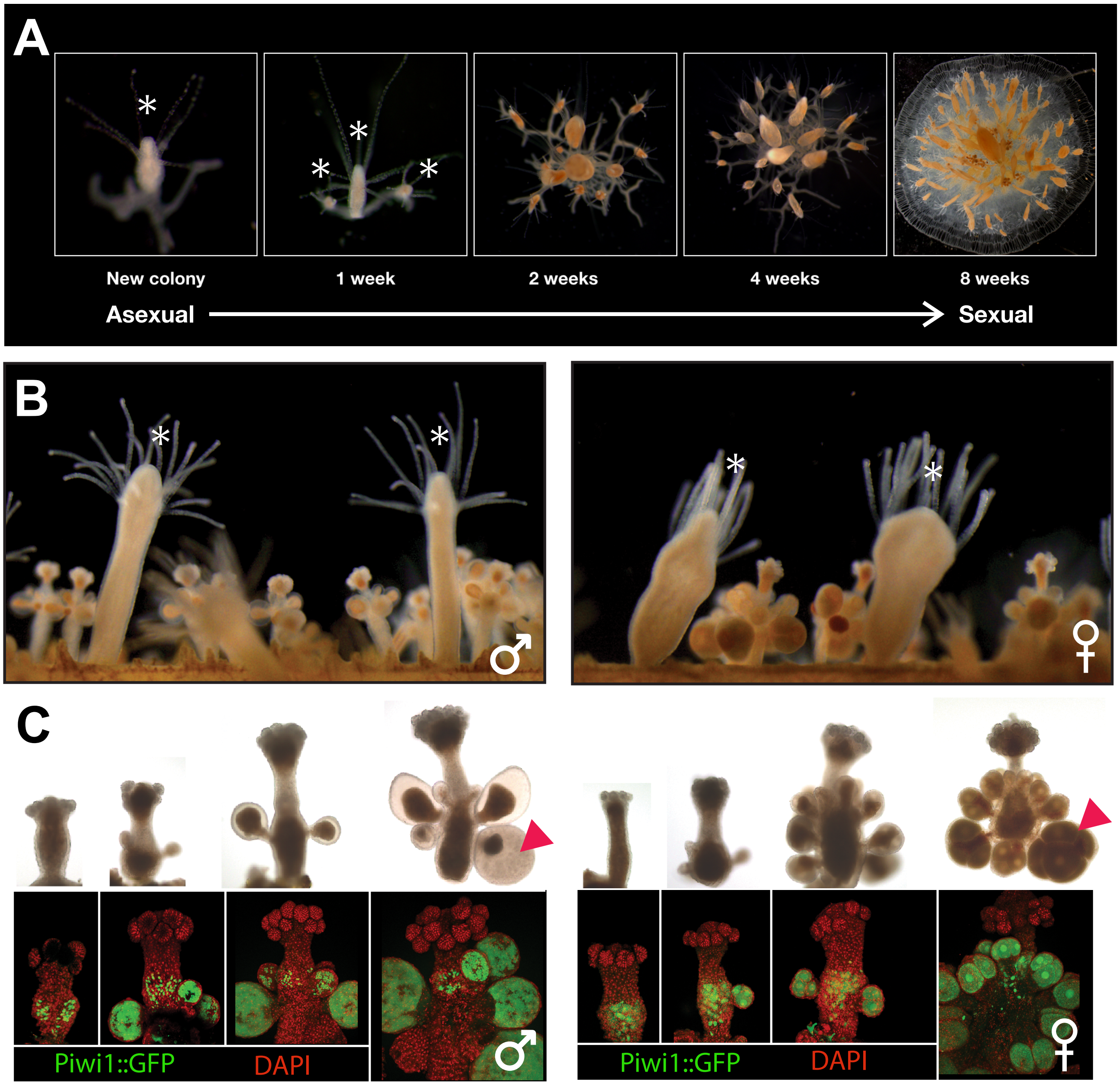


Fig. S1 – Adult *Hydractinia* asexual and sexual development. (A) Timeline of adult development from a primary polyp (left), clonal asexual growth and finally a sexually mature adult (right). (B) Male and female mature adult with feeding and sexual polyps. The mouth of the feeding polyp is indicated with an asterisk. (C) Stages of male and female sexual polyp formation. Lower panels are taken from a live *Piwi1* transgenic reporter that expresses GFP (green). Stages are counterstained with DAPI (red). The green Piwi+ cells are developing sperm and eggs within the gastroderm of the sexual polyps. Red arrowheads indicate mature gametes.

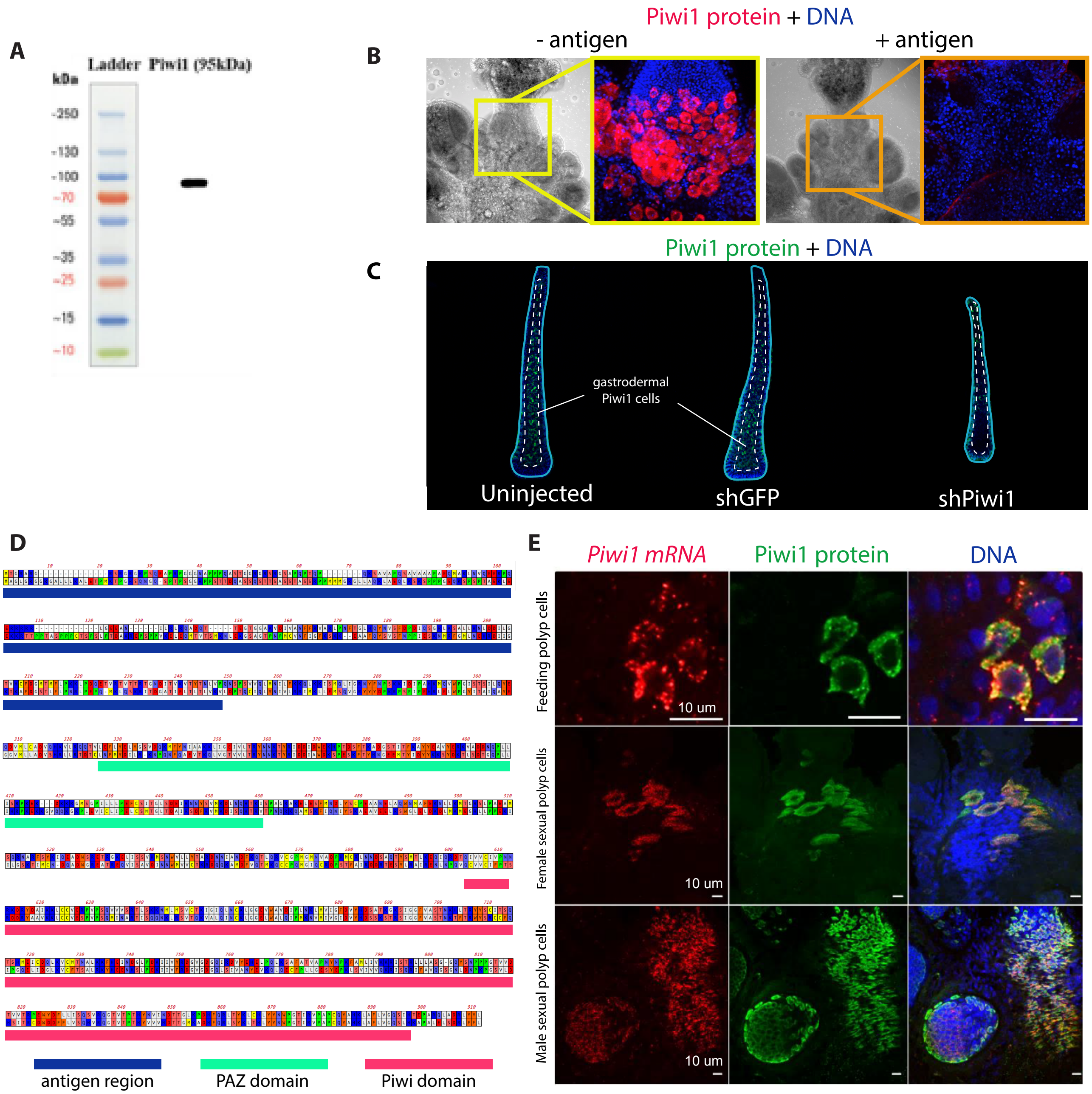


Fig. S2 – Characterization of anti-Piwi1 antibody. (A) Western blot of the rabbit polyclonal anti-Piwi1 one antibody from whole animal protein extract. (B) Piwi1 immunofluorescence in wild type female sexual polyps (Piwi1 antigen free, left; antigen competition, right). Yellow and orange squares represent corresponding area where the immunofluorescent image was taken. (C) Piwi1 shRNA-mediated knockdown abolishes the majority of Piwi1 positive cells identified with the gastroderm using anti-Piwi antibody. (D) Alignment of Piwi1 and Piwi2 protein shows a lack of conservation within the antigen region. (E) Colocalization of *Piwi1* mRNA (in situ hybridization) and Piwi1 protein (immunofluorescence) in diverse adult tissues. Blue staining is DAPI label in the DNA of cells.

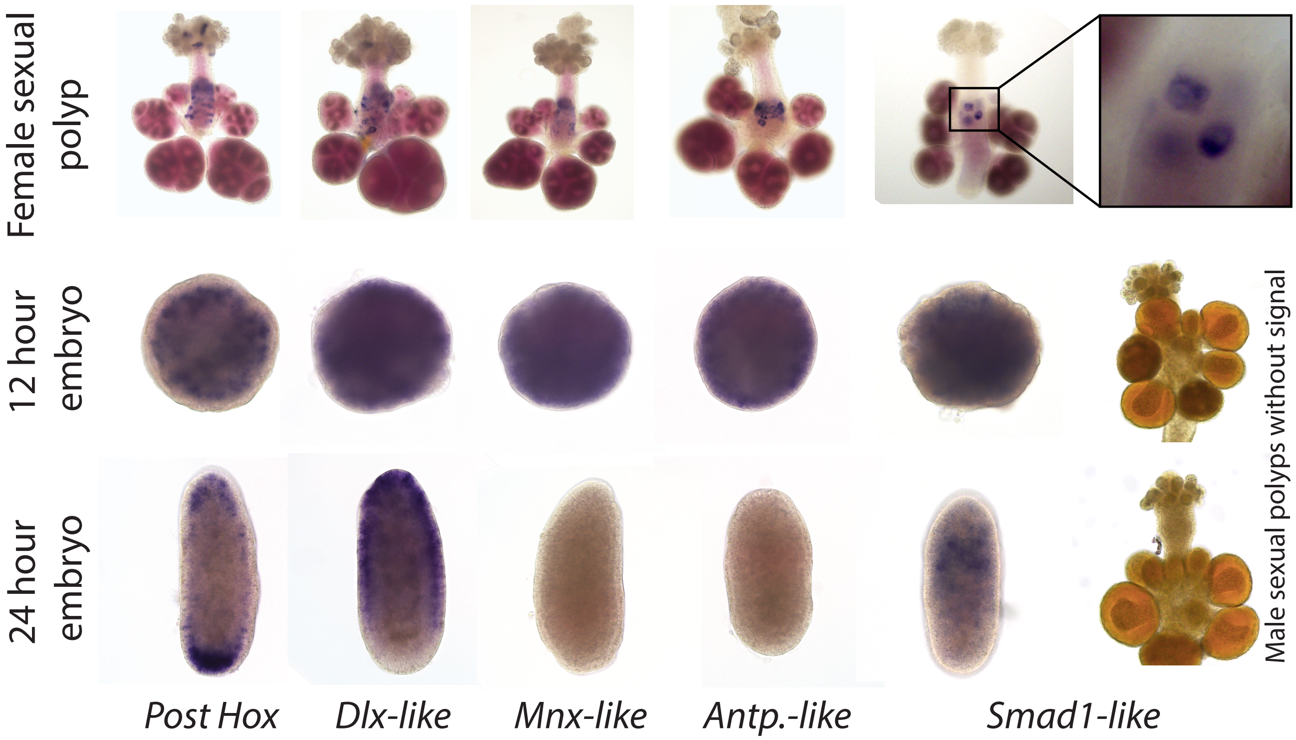


Fig. S3 – mRNA in situ hybridization of female specific gene expression in developing oocytes. Identified genes appear in developing oocytes and expression is also present in early stages of embryonic and larval development. These genes did not appear in males, indicating their potential involvement in maternal loading, based on their continued expression in early embryos.

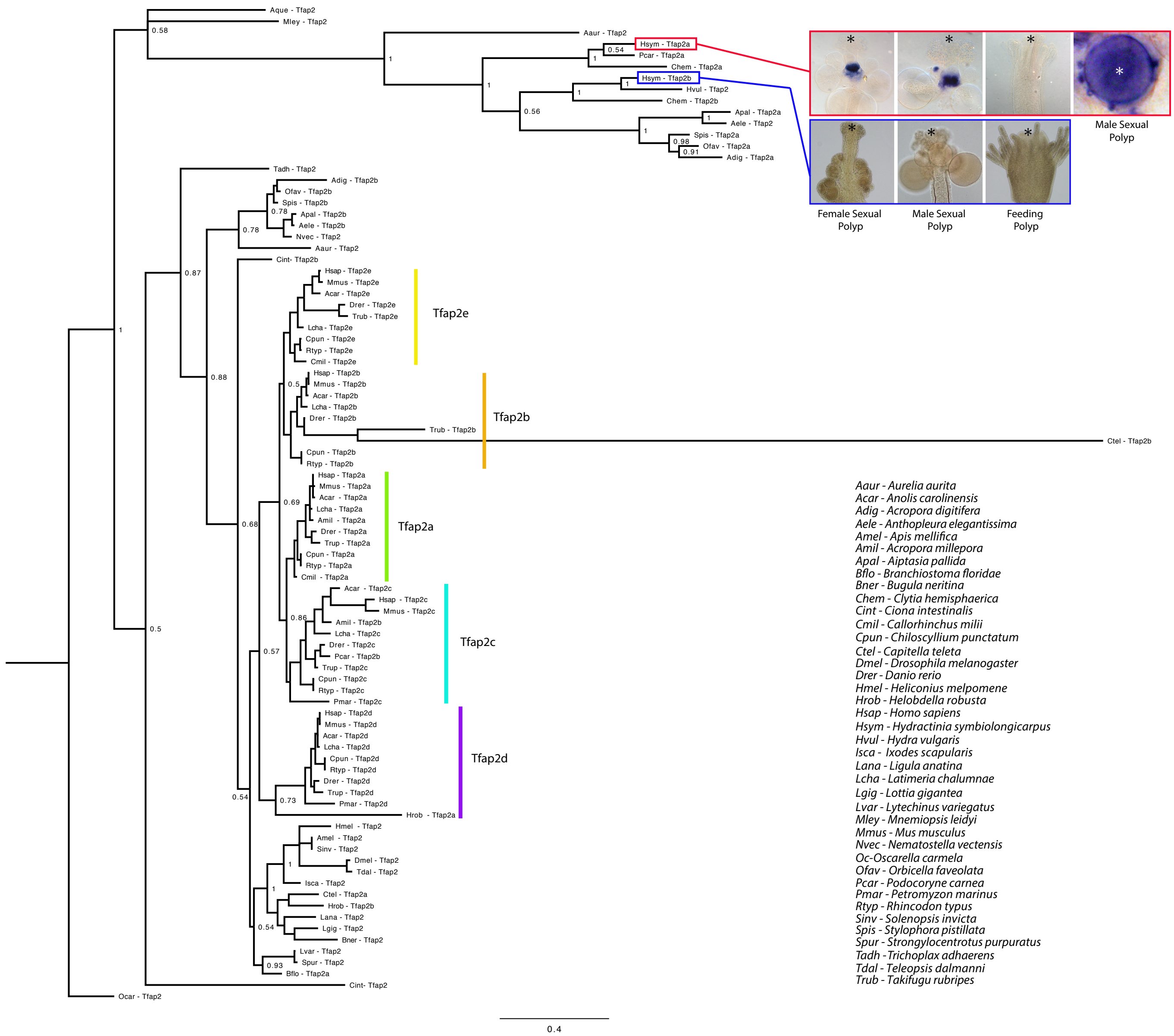


Fig. S4 – Phylogenetic tree of Tfap2 proteins in all major phylogenetic groups of animals. mRNA of *Hydractinia Tfap2a* and *Tfap2b* in sexual and feeding polyps. *Tfap2a* (*Tfap2*) mRNA is only detectable in the germinal zone of sexual polyps and can be found in both epidermal and gastrodermal cells. *Tfap2b* is undetectable. Asterisk indicates the oral side of the polyp.

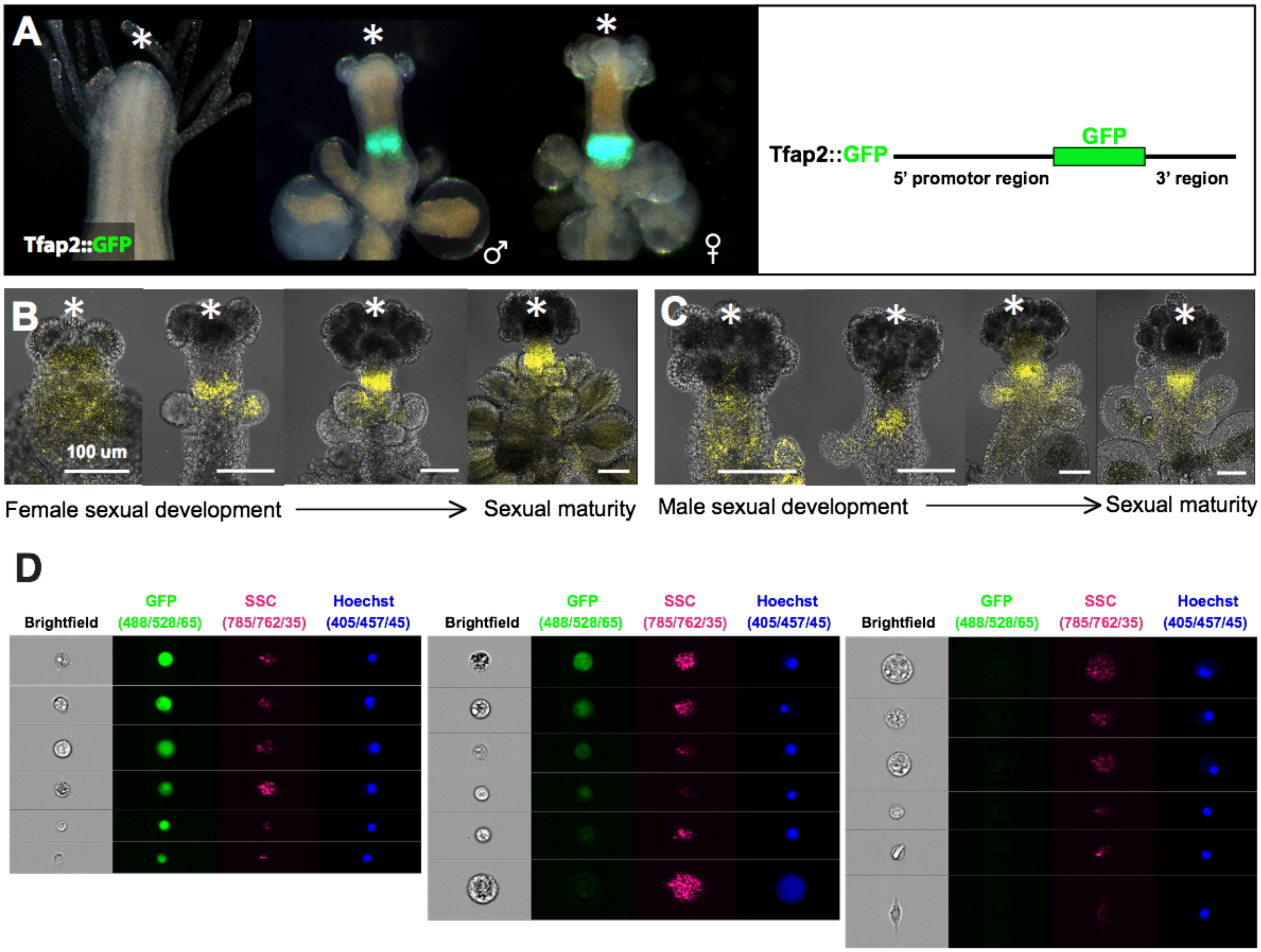


Fig. S5 – Tfap2 reporter and image stream analysis from live animals. (A) Live images of Tfap2::GFP in feeding (no GFP), male and female sexual polyps. The transgene consists of a 5' promoter region, the coding region of GFP, and the 3' region non-coding region of the *Tfap2* gene. Sequence information can be found in file S2. (B,C) GFP localization (yellow) is found throughout different stages of sexual development in male and female sexual polyps. (D) ImageStream analysis of GFP positive cells from disassociated female sexual polyps of the Tfap2::GFP reporter. Columns show GFP high, medium and low expressing cells.

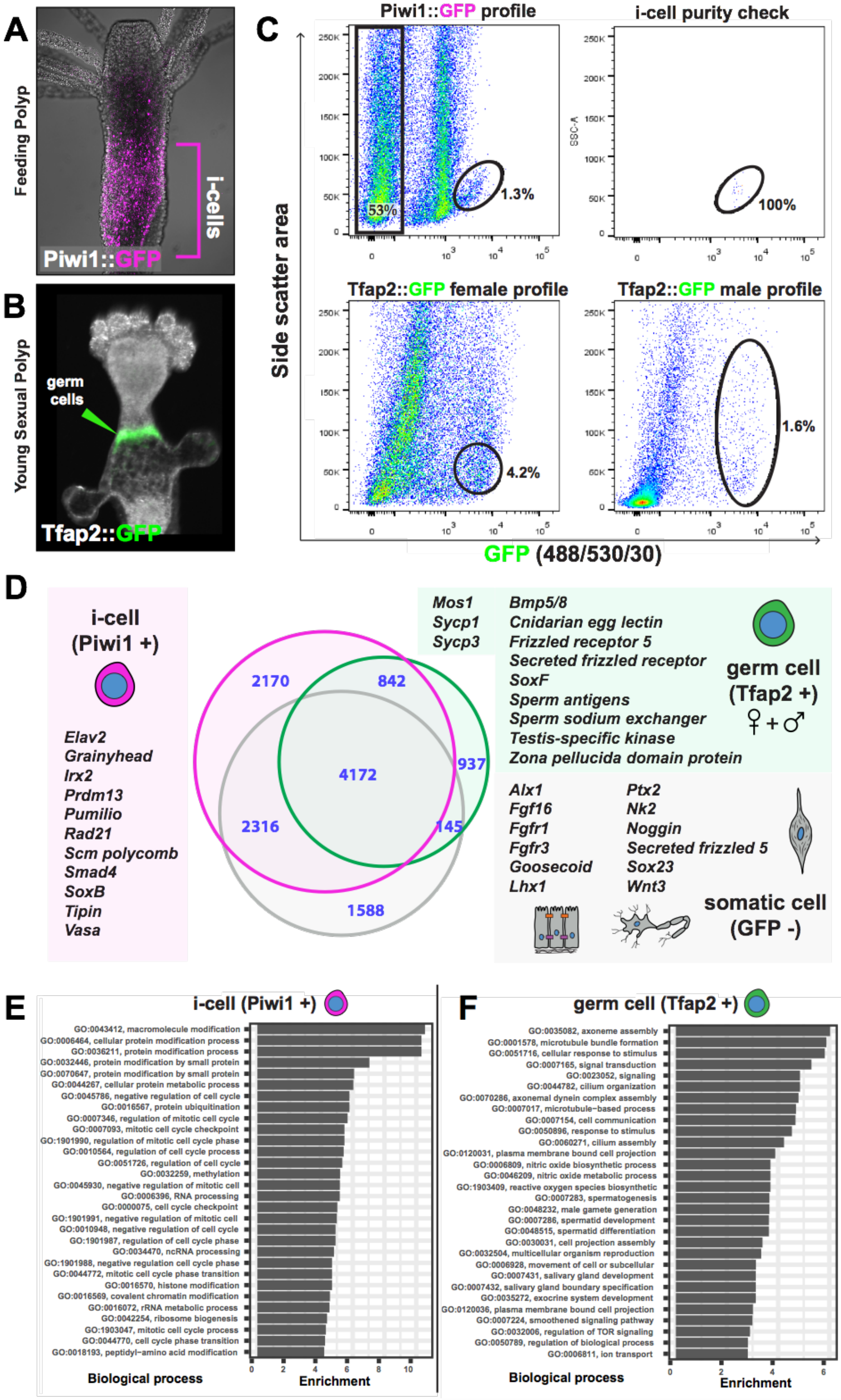

A
Feeding Polyp
i-cells
Piwi1::GFP
B
Young Sexual Polyp
germ cells
Tfap2::GFP
C
Piwi1::GFP profile
i-cell purity check
Tfap2::GFP female profile
Tfap2::GFP male profile
Side scatter area
GFP (488/530/30)
53%
1.3%
100%
4.2%
1.6%
D
i-cell (Piwi1 +)
Elav2
Grainyhead
Irx2
Prdm13
Pumilio
Rad21
Scm polycomb
Smad4
SoxB
Tipin
Vasa
2170
842
4172
937
2316
145
1588
Mos1
Sycp1
Sycp3
Bmp5/8
Cnidarian egg lectin
Frizzled receptor 5
Secreted frizzled receptor
SoxF
Sperm antigens
Sperm sodium exchanger
Testis-specific kinase
Zona pellucida domain protein
germ cell (Tfap2 +)
♀+♂
Alx1
Fgf16
Fgfr1
Fgfr3
Goosecoid
Lhx1
Ptx2
Nk2
Noggin
Secreted frizzled 5
Sox23
Wnt3
somatic cell (GFP -)
E
i-cell (Piwi1 +)
Biological process
Enrichment
F
germ cell (Tfap2 +)
Biological process
Enrichment

Fig. S6 - Transcriptomic analysis of i-cells, germ cells, and somatic cells. (A) Live image of a *Piwi1* transgenic reporter feeding polyp. (B) Live image of a *Tfap2* transgenic reporter sexual polyp. (C) i-cells, somatic cells, and germ cells sorting plots, and i-cell purity control. (D) Venn diagram of differentially expressed genes in i-cells, germ cells, and somatic cells. (E-F) GO-term enrichment in i-cells and germ cells.

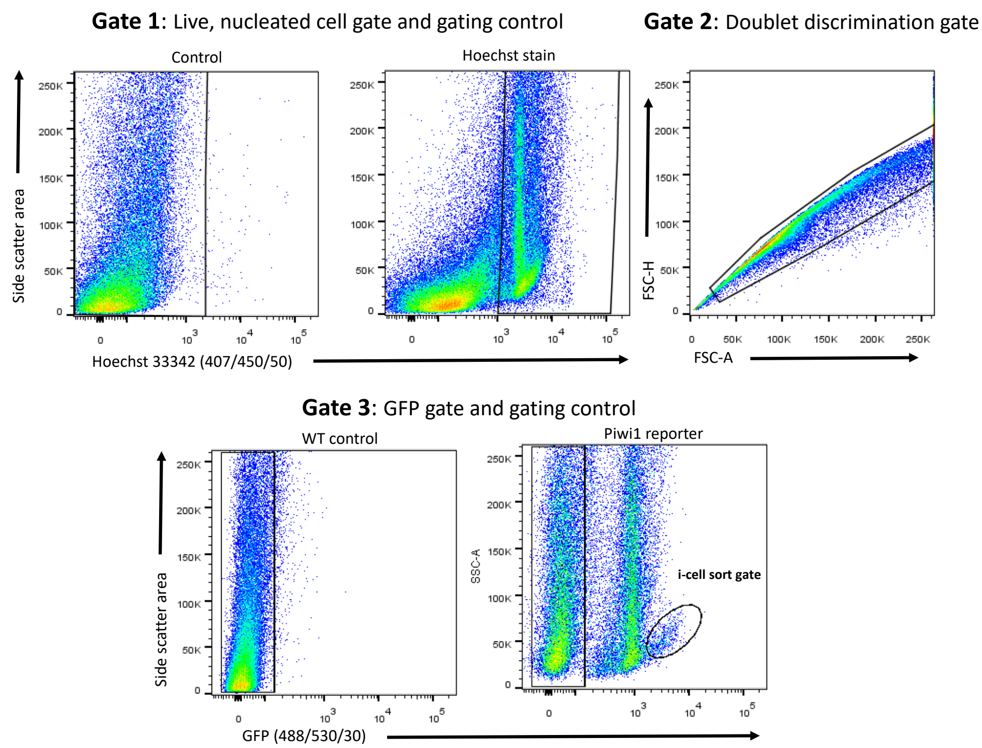
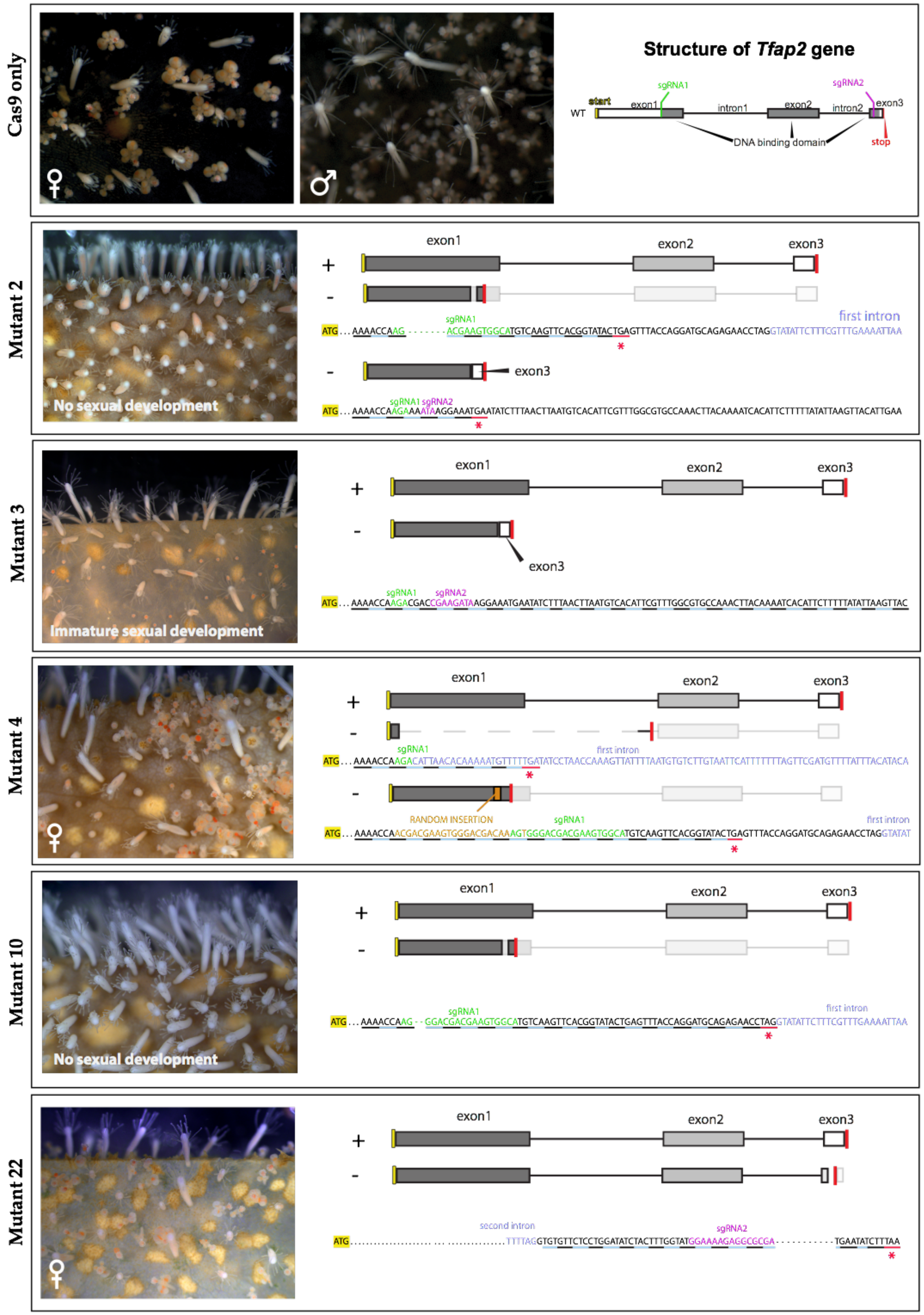


Fig. S7 - Representative gating strategy for sorting of GFP⁺ transgenic reporter cells. Live, nucleated cells were identified using Hoescht 33342 nucleic acid stain. Hoescht 33342 positively stained cells were gated using unstained animals as a gating control. Single cells were identified by doublet discrimination gating based on forward scatter height versus forward scatter area parameters. GFP expressing cells were gated using wild type animals as a gating control. Depicted here is the gating strategy and i-cell sort gate specific for the *Piwi1* reporter animal.



Cas9 only
♀
♂
Structure of Tfap2 gene
WT
start
exon1
sgRNA1
intron1
exon2
intron2
sgRNA2
exon3
DNA binding domain
stop
Mutant 2
No sexual development
exon1
exon2
exon3
+
-
first intron
sgRNA1
sgRNA2
Mutant 3
Immature sexual development
Mutant 4
RANDOM INSERTION
Mutant 10
Mutant 22
second intron

Fig. S8 – Characterization of CRISPR/Cas9 generated Tfap2 mutants. The *Tfap2* gene consists of three exons and a 3' DNA binding domain that spans the three exons. Generated mutants all maintained a wild type allele (+) and each mutant had different variations of cutting between the sgRNA1 and sgRNA2. In mutant 2, we found a single cut and frameshift mutation. We also identified a double cut resulting in re-sectioning of the gene and a loss of the two introns and exon two. These animals had no sexual development. Mutant 3 had a double cut and loss of exon two. These animals initiate sexual polyp formation without maturation. Mutant 4 had two different single cut events (both at the sgRNA1). One event results in a loss of the first intron and a premature stop codon. The second event created an insertion mutation and a premature stop codon. This genotype was selected for when making our homozygous animals. This animal could still undergo sexual reproduction and offspring carry these mutations. Mutant 10 exhibited a single cut event (sgRNA1) and resulted in a small deletion and premature stop codon. These animals never showed signs of sexual development. Mutant 22 exhibited a single cut event (sgRNA2) resulting in animals that have a small truncated 3'end but still underwent sexual reproduction.

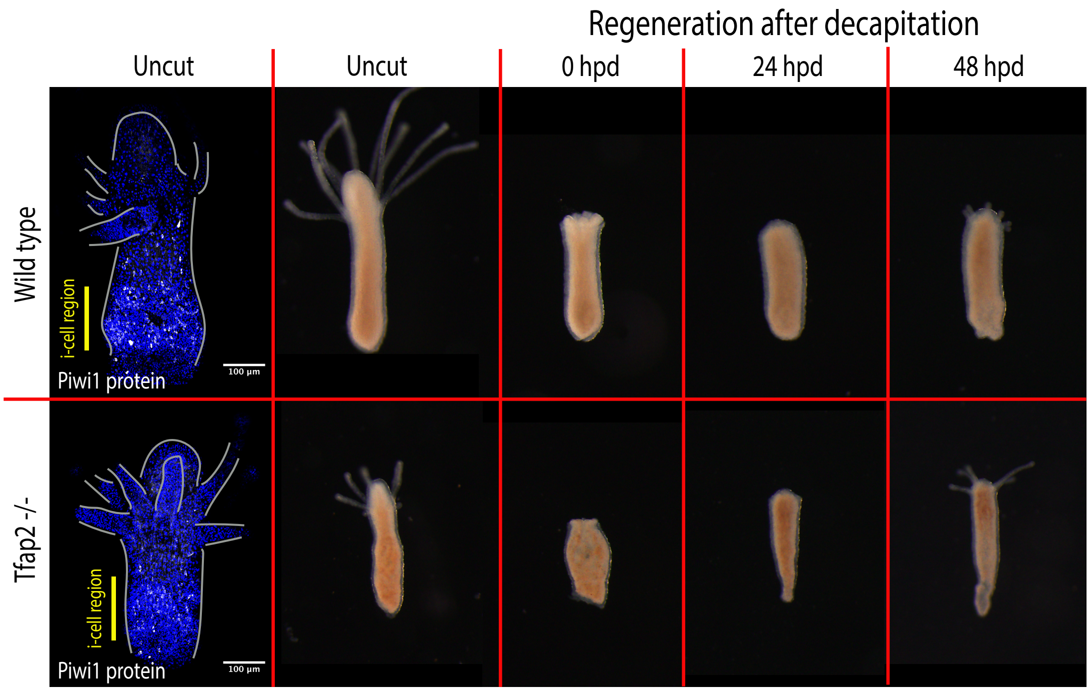
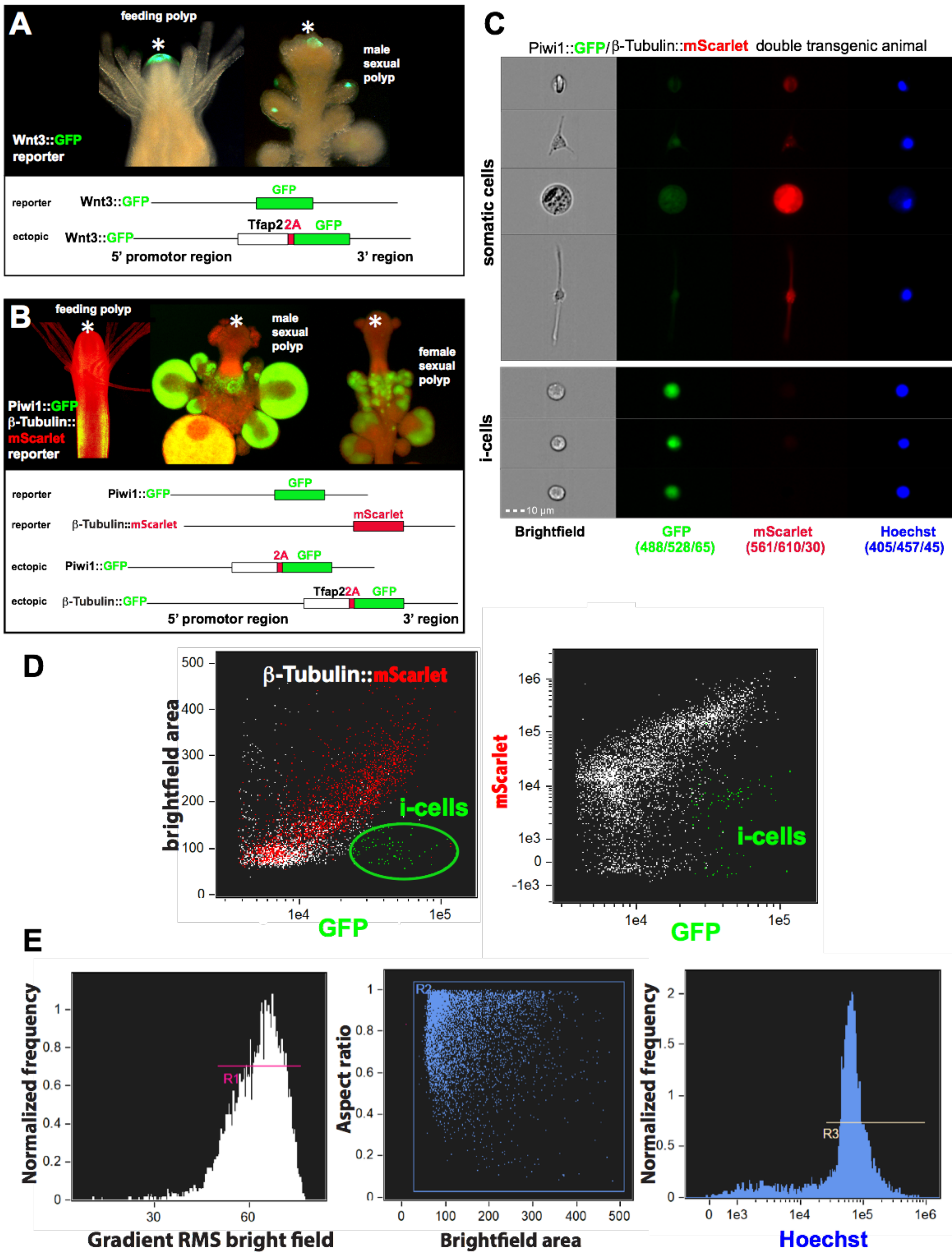


Fig. S9 – Characterization of a homozygous *Tfap2* mutant. Neither the distribution of i-cells (Piwi1 protein localization) nor its regenerative potential appear compromised by the loss of *Tfap2*.



A
feeding polyp
male sexual polyp
Wnt3::GFP reporter
reporter Wnt3::GFP
GFP
ectopic Wnt3::GFP
Tfap2 2A GFP
5' promotor region
3' region
B
feeding polyp
male sexual polyp
female sexual polyp
Piwi1::GFP β-Tubulin::mScarlet reporter
reporter Piwi1::GFP
reporter β-Tubulin::mScarlet
mScarlet
ectopic Piwi1::GFP
2A GFP
ectopic β-Tubulin::GFP
Tfap2 2A GFP
5' promotor region
3' region
C
Piwi1::GFP/β-Tubulin::mScarlet double transgenic animal
somatic cells
i-cells
10 µm
Brightfield
GFP (488/528/65)
mScarlet (561/610/30)
Hoechst (405/457/45)
D
β-Tubulin::mScarlet
brightfield area
i-cells
GFP
mScarlet
i-cells
GFP
E
Normalized frequency
R1
Gradient RMS bright field
Aspect ratio
R2
Brightfield area
Normalized frequency
R3
Hoechst

Fig. S10 - Characterization of transgenic reporter constructs used to ectopically express Tfap2. (A) Wnt3 reporter, expressing GFP at the oral ends of feeding and sexual polyps, and at the sporosacs' tips in sexual polyps. The structure of the reporter and Tfap2 ectopic expression constructs is shown. (B) Piwi1/beta-tubulin double transgenic reporter animal, expressing GFP in i-cells and germ cells, and mScarlet in all somatic cells. The expression patterns of the fluorescent proteins and the constructs are shown. (C) Imaging flow cytometry images of dissociated Piwi1/beta-tubulin double transgenic cells from whole polyps. beta-tubulin promoter-driven mScarlet is not expressed in i-cells. Somatic cells show low GFP fluorescence, probably representing residual Piwi1::GFP expression in their i-cell progenitors, which is downregulated following i-cell differentiation. (D) Imaging flow cytometry plots of whole polyp dissociated cells shown in (C). i-cells do not express the beta-tubulin::mScarlet transgene. (E) Gating strategy of the above analyses.

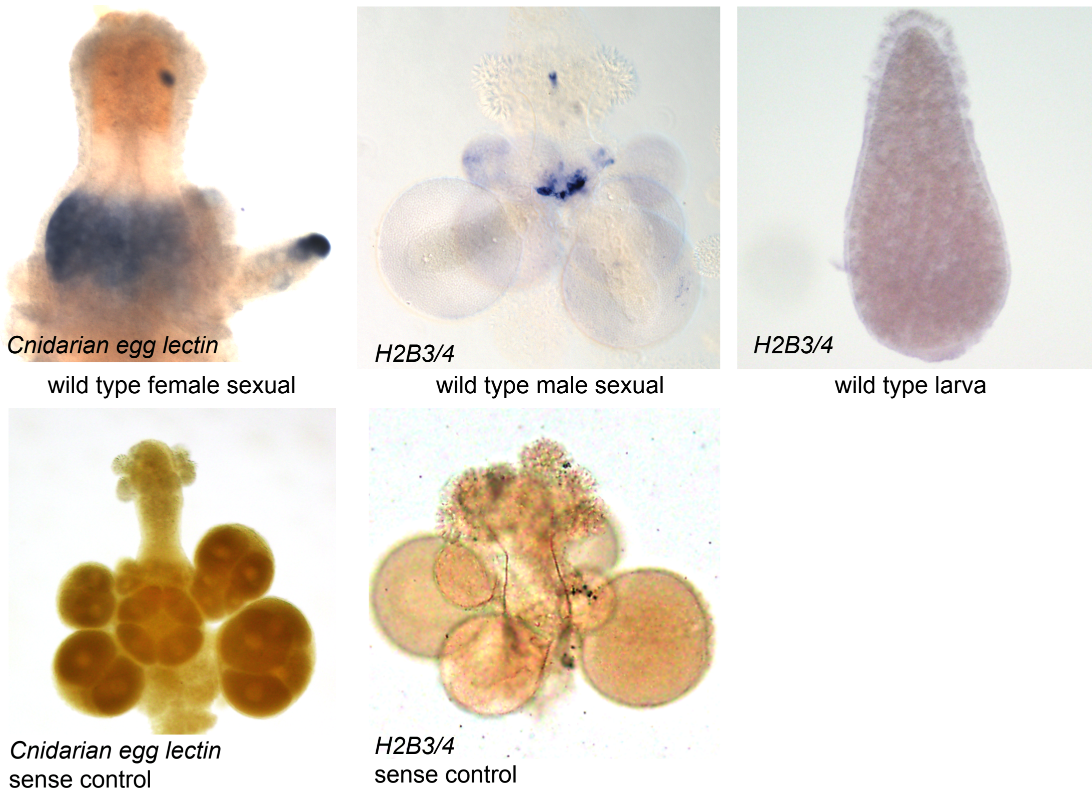


Fig. S11 – Male and female germ cell specific genes are expressed near the neck of the sexual polyp. *Cnidarian egg lectin (Cel)* is expressed in female germ cells. *H2B3/4* is expressed in male germ cells and not in wild type larvae.

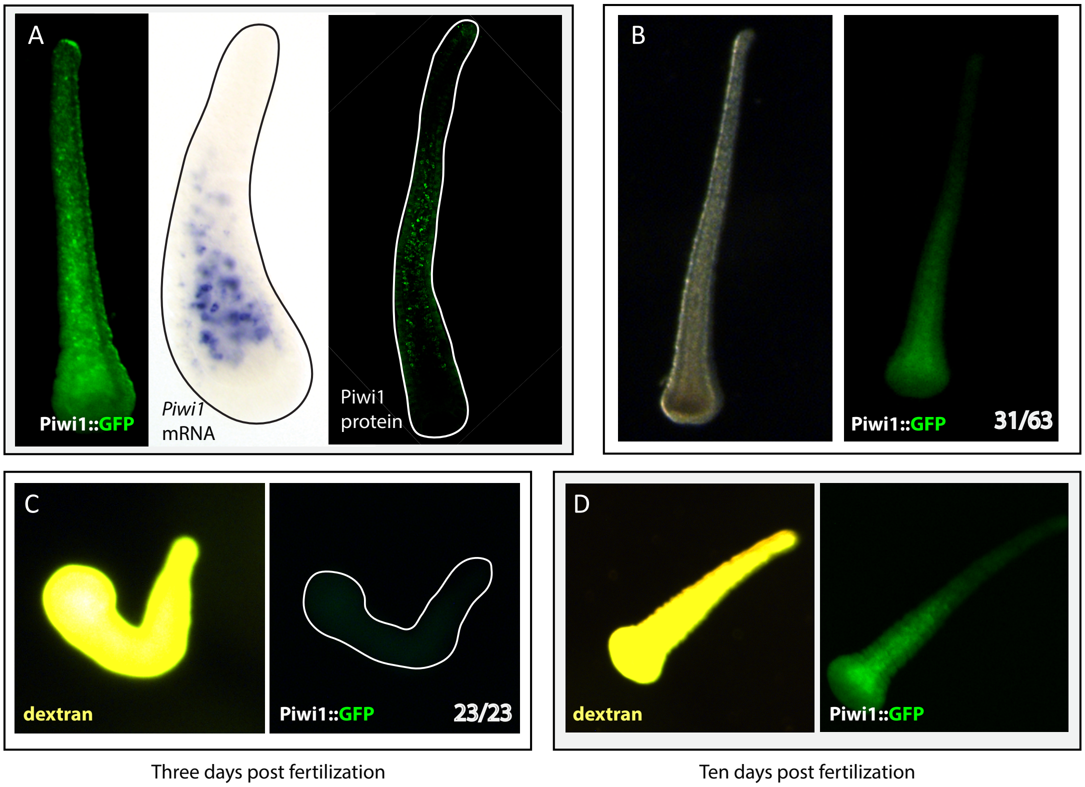


Fig. S12 - Efficacy of shRNA targeting GFP in a *Piwi1* transgenic reporter larvae that was injected at the zygote stage. (a) Normal expression of Piwi1, viewed in an uninjected transgenic reporter larva, and mRNA in situ hybridization and protein immunofluorescence, respectively. (b) Uninjected larvae resulting from crossing a heterozygote transgenic *Piwi1* reporter male with a wild type female. About 50% of the offspring carry the transgene as expected. (c, d) The same pool of larvae after injecting them with the GFP-targeting shRNA. GFP is downregulated in 100% of the larvae 3 days post injection but is upregulated starting 10 days post injection. Dextran-TRIC was used as injection tracer.

Table S1 - Differentially expressed genes between heads and bodies of feeding and sexual polyps. Data are organized into different tabs within the file to delineate tissue specific differential expression between tissues. All values are sorted based on fold change, and genes are described by their top blast hit. This dataset is summarized in Fig. 2.

File S1 - Differentially expressed genes and GO term enrichment between somatic cells, i-cells, and germ cells. Dataset is the accumulated results from FAC sorted cells from Tfap2, Piwi1 transgenic reporters. This dataset is summarized in fig. S6.

File S2 – DNA sequences of all plasmids used within this study. Each construct contains 5' promoter elements, the coding region of the fluorescent protein of fluorescent protein Tfap2 fusion, and is flanked by the 3' non-coding region of each gene.

Movie S1 - Live image of transgenic Tfap2 reporter animal. Green region is the germinal zone and is GFP positive in Tfap2::GFP animals. This animal is a female sexual polyp, with mature eggs forming near the bottom of the animal. Timeseries taken over the course of twenty-four hours. Series shows that germ cell production appears to be a slow process (low turn-over).

Movie S2 - Live image of migrating Tfap2 wild type cells that carry a reporter transgene in the tissue of a non-fluorescent Tfap2 mutant. This movie shows the transgenic donor animal (bottom) that has wild type Tfap2 alleles. Cells from the donor and can be seen migrating into the Tfap2 mutant tissue (top).

References

1. C. G. Extavour, M. Akam, Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869-5884 (2003).
2. L. W. Buss, Slime molds, ascidians, and the utility of evolutionary theory. *Proc. Natl. Acad. Sci. USA* **96**, 8801-8803 (1999).
3. N. W. Blackstone, B. D. Jasker, Phylogenetic considerations of clonality, coloniality, and mode of germline development in animals. *J Exp Zoolog Part B Mol Dev Evol* **297**, 35-47 (2003).
4. A. L. Radzvilavicius, Z. Hadjivasiliou, A. Pomiankowski, N. Lane, Selection for Mitochondrial Quality Drives Evolution of the Germline. *PLOS Biology* **14**, e2000410 (2016).
5. C. E. Juliano, G. M. Wessel, Versatile germline genes. *Science* **329**, 640-641 (2010).
6. T. Nakamura, C. G. Extavour, The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket Gryllus bimaculatus. *Development* **143**, 255-263 (2016).
7. N. Irie *et al.*, SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell* **160**, 253-268 (2015).
8. E. Magnusdottir *et al.*, A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol*, (2013).
9. M. Saitou, S. C. Barton, M. A. Surani, A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300 (2002).
10. J. M. Gahan, B. Bradshaw, H. Flici, U. Frank, The interstitial stem cells in *Hydractinia* and their role in regeneration. *Curr Opin Genet Dev* **40**, 65-73 (2016).
11. T. C. Bosch, C. N. David, Stem cells of *Hydra magnipapillata* can differentiate into somatic cells and germ line cells. *Dev Biol* **121**, 182-191 (1987).
12. A. Weismann, *Die Entstehung der Sexualzellen bei Hydromedusen.* (Gustav Fischer, Jena, 1883).
13. W. A. Müller, Experimentele Untersuchungen über Stockentwicklung, Polypendifferenzierung und Sexualchimären bei *Hydractinia echinata*. *Roux' Arch. für Entwicklungsmechanik* **155**, 181-268 (1964).
14. C. E. Juliano, S. Z. Swartz, G. M. Wessel, A conserved germline multipotency program. *Development* **137**, 4113-4126 (2010).
15. B. Bradshaw, K. Thompson, U. Frank, Distinct mechanisms underlie oral vs aboral regeneration in the cnidarian *Hydractinia echinata*. *eLife* **4**, (2015).
16. S. M. Sanders, M. Shcheglovitova, P. Cartwright, Differential gene expression between functionally specialized polyps of the colonial hydrozoan Hydractinia symbiolongicarpus (Phylum Cnidaria). *BMC genomics* **15**, 406 (2014).
17. A. M. Villeneuve, K. J. Hillers, Whence meiosis? *Cell* **106**, 647-650 (2001).
18. M. A. Ramesh, S. B. Malik, J. M. Logsdon, Jr., A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Curr Biol* **15**, 185-191 (2005).
19. W. A. Pastor *et al.*, TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nature Cell Biology* **20**, 553-564 (2018).
20. F. Nakaki *et al.*, Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222-226 (2013).

21. K. Sasaki *et al.*, Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* **17**, 178-194 (2015).
22. S. Aramaki *et al.*, A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev Cell* **27**, 516-529 (2013).
23. S. Weber *et al.*, Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biology of reproduction* **82**, 214-223 (2010).
24. B. Ewen-Campen, S. Donoughe, D. N. Clarke, C. G. Extavour, Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect. *Curr Biol*, (2013).
25. A. Ikmi, S. A. McKinney, K. M. Delventhal, M. C. Gibson, TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan Nematostella vectensis. *Nat Commun* **5**, 5486 (2014).
26. J. M. Gahan *et al.*, Functional studies on the role of Notch signaling in *Hydractinia* development. *Dev Biol* **428**, 224-231 (2017).
27. T. Q. DuBuc, T. B. Stephenson, A. Q. Rock, M. Q. Martindale, Hox and Wnt pattern the primary body axis of an anthozoan cnidarian before gastrulation. *Nature Communications* **9**, 2007 (2018).
28. S. M. Sanders *et al.*, CRISPR/Cas9-mediated gene knockin in the hydroid Hydractinia symbiolongicarpus. *BMC genomics* **19**, 649 (2018).
29. R. D. Rosengarten, M. L. Nicotra, Model Systems of Invertebrate Allorecognition. *Curr Biol* **21**, R82-R92 (2011).
30. T. Künzel *et al.*, Migration and differentiation potential of stem cells in the cnidarian *Hydractinia* analyzed in GFP-transgenic animals and chimeras. *Dev Biol* **348**, 120-129 (2010).
31. C. Rios-Rojas, J. Bowles, P. Koopman, On the role of germ cells in mammalian gonad development: quiet passengers or back-seat drivers? *Reproduction (Cambridge, England)* **149**, R181-191 (2015).
32. Z. Cao, X. Mao, L. Luo, Germline Stem Cells Drive Ovary Regeneration in Zebrafish. *Cell reports* **26**, 1709-1717.e1703 (2019).
33. G. Plickert, V. Jacoby, U. Frank, W. A. Müller, O. Mokady, Wnt signaling in hydroid development: Formation of the primary body axis in embryogenesis and its subsequent patterning. *Developmental Biology* **298**, 368-378 (2006).
34. D. J. Duffy, G. Plickert, T. Künzel, W. Tilmann, U. Frank, Wnt signaling promotes oral but suppresses aboral structures in *Hydractinia* metamorphosis and regeneration. *Development* **137**, 3057-3066 (2010).
35. K. Hensel, T. Lotan, S. M. Sanders, P. Cartwright, U. Frank, Lineage-specific evolution of cnidarian Wnt ligands. *Evol Dev*, (2014).
36. A. Torok *et al.*, The cnidarian Hydractinia echinata employs canonical and highly adapted histones to pack its DNA. *Epigenetics Chromatin* **9**, 36 (2016).
37. S. He *et al.*, An axial Hox code controls tissue segmentation and body patterning in Nematostella vectensis. *Science* **361**, 1377-1380 (2018).
38. B. Mali, R. C. Millane, G. Plickert, M. Frohme, U. Frank, A polymorphic, thrombospondin domain-containing lectin is an oocyte marker in *Hydractinia*: implications for germ cell specification and sex determination. *Int J Dev Biol* **55**, 103-108 (2011).
39. T. Kobayashi, M. A. Surani, On the origin of the human germline. *Development* **145**, (2018).

40. F. Fang *et al.*, A PAX5–OCT4–PRDM1 developmental switch specifies human primordial germ cells. *Nature Cell Biology* **20**, 655-665 (2018).
41. D. S. Bindels *et al.*, mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature methods* **14**, 53-56 (2016).
42. W. A. Müller, G. Buchal, Metamorphoseinduktion bei Planulalarven. II. Induktion durch monovalente Kationen: Die Bedeutung des Gibbs-Donnan Verhältnisses und der Ka+Na+-ATPase. *Wilhelm Roux Arch.* **173**, 122-135 (1973).
43. Y. J. Liew, M. Aranda, C. R. Voolstra, Reefgenomics.Org - a repository for marine genomics data. *Database : the journal of biological databases and curation* **2016**, (2016).
44. Y. Hara *et al.*, Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nature Ecology & Evolution* **2**, 1761-1771 (2018).
45. L. Leclere *et al.*, The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. *Nat Ecol Evol* **3**, 801-810 (2019).
46. K. Katoh, J. Rozewicki, K. D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*, (2017).
47. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
48. A. Rambaut. (2007).
49. H. Flici *et al.*, An evolutionarily conserved SoxB-Hdac2 crosstalk regulates neurogenesis in a cnidarian. *Cell reports* **18**, 1395-1409 (2017).
50. C. E. Juliano *et al.*, PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc Natl Acad Sci U S A* **111**, 337-342 (2014).
51. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, (2014).
52. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **35**, 1026-1028 (2017).
53. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644-652 (2011).
54. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360 (2015).
55. D. Gilbert, Gene-omes built from mRNA seq not genome DNA. *7th annual arthropod genomics symposium, Notre Dame.*, (2013).
56. P. Toronen, A. Medlar, L. Holm, PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res* **46**, W84-W88 (2018).
57. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
58. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology* **32**, 896 (2014).
59. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, (2010).
60. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).
61. D. Yu, W. Huber, O. Vitek, Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics* **29**, 1275-1282 (2013).

62. D. Hebenstreit *et al.*, RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular systems biology* **7**, 497 (2011).
63. T. Benaglia, D. Chauveau, D. Hunter, D. Young, mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software* **32**, 1-29 (2009).